Web Appendix for Probability of Fit Failure with Reuse of N95 Respirators

Bruno Maranhao, MD PhD Alex W Scott, MD Jooyoung Maeng, MD
Ziyan Song, MD Ramya Baddigam, MD C Ryan King, MD PhD
Molly McCormick, MD Ivan Kangrga, MD PhD Ryan Guffey, MD

June 08, 2020

1 Modeling assumptions and strategy

The sample is generated cross-sectionally, but can also be thought of as survival data. Each fit test provides interval censored information: masks observed to fail a fit test at time t failed at some point in $(0, t)$ (masks are assumed to have been initially functional). Masks that pass the fit test fail some time in the future $(t, t_{\max})$. For survival analysis we assume that each mask is independent with failure hazard from from a common distribution. Our parametric survival model fits a 3-parameter generalized gamma hazard to the interval censored data.

Non-parametric or semi-parametric estimators which do not assume a functional form of the hazard are common in traditional survival analysis, and are available for interval censored data as well. However, non-parametric estimates (NPMLE) for interval censored data are not unique; in some regions any baseline hazard matching continuity requirements is equally supported by the data. Confidence intervals for interval censored baseline hazards are an active area of research (Wang et al. 2016; Zhao et al. 2019), with completely non-parametric estimation methods yielding slow $n^{\frac{1}{3}}$ convergence without a known asymptotic law (Groeneboom and Wellner 1992). Several semi-parametric methods have recently been shown to have $n^{\frac{1}{2}}$ convergence with Gaussian asymptotics making them amenable to bootstrapping (Wang et al. 2016; Bouaziz, Lauridsen, and Nuel 2018; Zhao et al. 2019; Li, Pak, and Todem 2019). Proportional hazards regression was used to model factors other than duration of use in this framework.

Modeling the data as cross-sectional observations of fit passing versus failure as a function of time is also supportable because no participant-mask pair is observed more than once and there is minimal information to support extrapolation from before the earliest test time. Because only a very small number of individuals are tested on more than one mask, we do not model or adjust for participant effects. Integrated b-splines and constrained step functions allow modeling the failure log-odds as a monotone increasing function of time in a traditional logistic regression analysis. This approach yields much better studied statistical behavior; however, given the modest sample size the probability statements produced should be taken as approximate. The cross-sectional model does not encode any behavior before the earliest test time (e.g. that all masks initially fit). We found that the cross-sectional model and non-parametric and semi-parametric survival models yield essentially identical point estimates when tuned appropriately, which is anticipated based on the use of step functions and splines to model the baseline hazard in the survival routines.

Factors beyond use duration were analyzed with proportional hazards regression (interval survival) or logistic regression (cross sectional). Calculations were performed in R 3.6.3 with package flexsurv (Jackson 2016) for parametric survival modeling, icenReg (Anderson-Bergman 2017) for non-parametric survival modeling, and cgam (Liao and Meyer 2019) for monotone logistic regression. Reported confidence intervals are at the 0.95 level. Because this was an exploratory study, no pre-determined p-value threshold was applied. Code and data to recreate this report can be found at https://github.com/cryanking/n95_refit

Table 1: Descriptive statistics stratified by fit test failure. P-values for quantitative variables by Mann-Whitney U, factor variables by Fisher's exact test.

	level	Pass	Fail	p
n		37	37	
Days Worn (median [IQR])		7 [5, 12]	8 [5, 12]	0.522
Sterilizations (median [IQR])		1 [0, 1]	1 [0, 2]	0.123
Uses per day (mean (SD))		3.7 (3.1)	2.8 (1.4)	0.141
Uses (median [IQR])		20 [15, 40]	18 [12, 35]	0.547
Sex (%)	Female	17 (45.9)	29 (78.4)	0.008
	Male	20 (54.1)	8 (21.6)	
Title (%)	Attending	10 (27.0)	4 (10.8)	0.088
	CRNA	16 (43.2)	26 (70.3)	
	Fellow	1 (2.7)	1 (2.7)	
	Resident	10 (27.0)	6 (16.2)	
Respirator (%)	3M 1804 Vflex	31 (83.8)	33 (89.2)	0.736
	3m 1860 Teal	6 (16.2)	4 (10.8)	
Fits well (%)	No	2 (5.4)	10 (27.0)	0.024
	Yes	35 (94.6)	27 (73.0)	
Mask quality (%)	Good	11 (29.7)	23 (62.2)	0.001
	Like New	25 (67.6)	10 (27.0)	
	Poor	1 (2.7)	4 (10.8)	

Table 2: Failure fractions by duration worn. Confidence intervals by Wilson's score method.

day_group	n	mean.failure	lower.conf	upper.conf
(0.5,5]	23	0.52	0.33	0.71
(5,8]	16	0.44	0.23	0.67
(8,12]	19	0.47	0.27	0.68
(12,61]	16	0.56	0.33	0.77

2 Results

Table 1 provides descriptive statistics. Notably, male participants were less likely to fail fit tests both marginally and adjusting for the number of days worn (OR = 0.7 95% CI 0.6 to 0.9). Figure 1 displays histograms of number of days worn, number of times used, and number of times sterilized. Surprisingly, the use duration and intensity was not markedly different between failed and passed mask fits. Breaking the data into groups by number of days worn, the failure fraction is displayed in Table 2. Table 2 shows that the failure fraction at all times is high; approximately 50% in all time categories.

Figure 2 displays mask survival as a function of days worn using cross-sectional approaches. A monotone fit (I-splines) and natural spline (no restrictions) can be seen to agree well with non-parametric monotone methods. The failure fraction at day 4 is 0.46 (95% CI 0.31, 0.62). At day 10 the failure has increased to 0.5 (95% CI 0.36, 0.63) and day 15 to 0.55 (95% CI 0.38, 0.71).

Figure 3 compares interval censored survival-based analyses and the cross-sectional analysis. The parametric model diverges from the non-parametric MLE at the extremes of time, as it requires a smoother change between the initial state (0 failure) to the early observed times. The generalized gamma and Weibull parametric families fit well. Figure 4 displays the results with gamma and weibull hazards.

We also computed a traditional Kaplan-Meier curve treating the data as right censored only (assuming that participants joined the experiment about when their masks failed). This is displayed in Figure 5, which shows much lower estimated early failure fractions as the later failures are no longer considered a possible

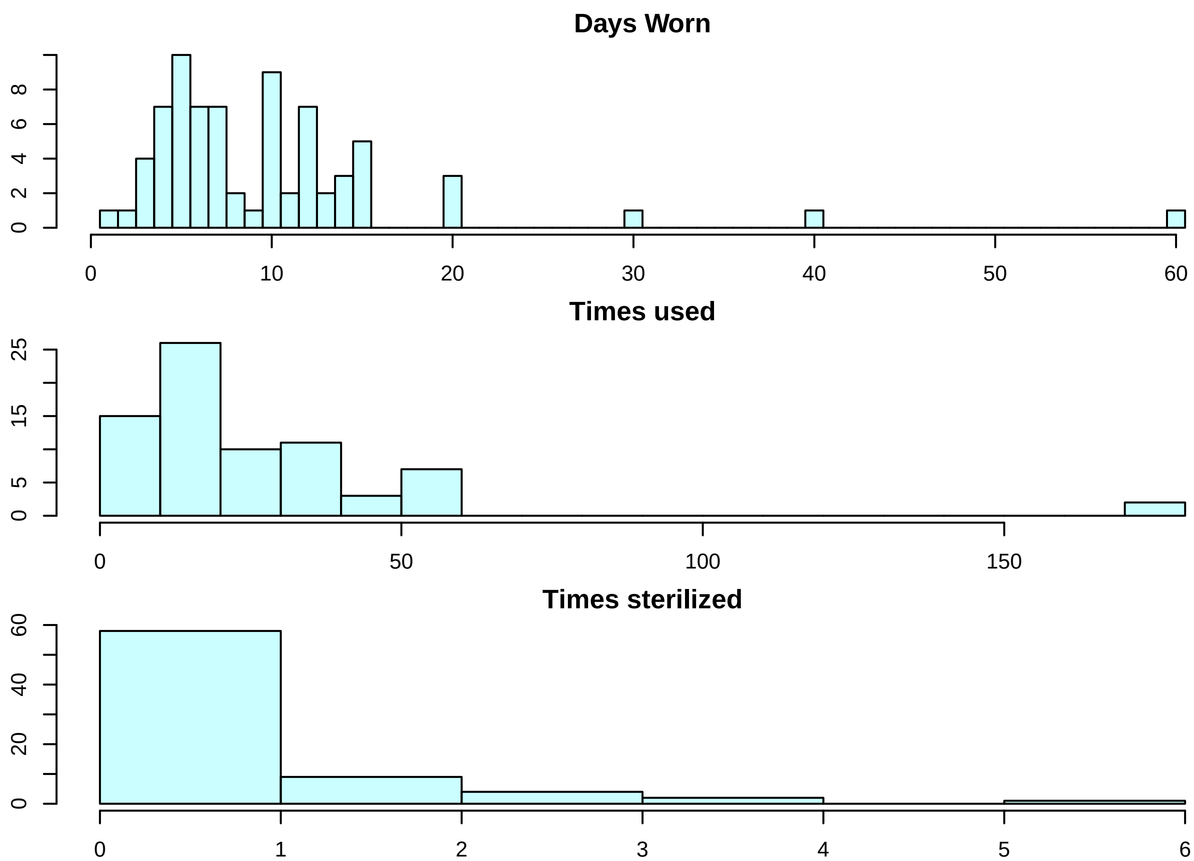

Figure 1: Mask use histograms.

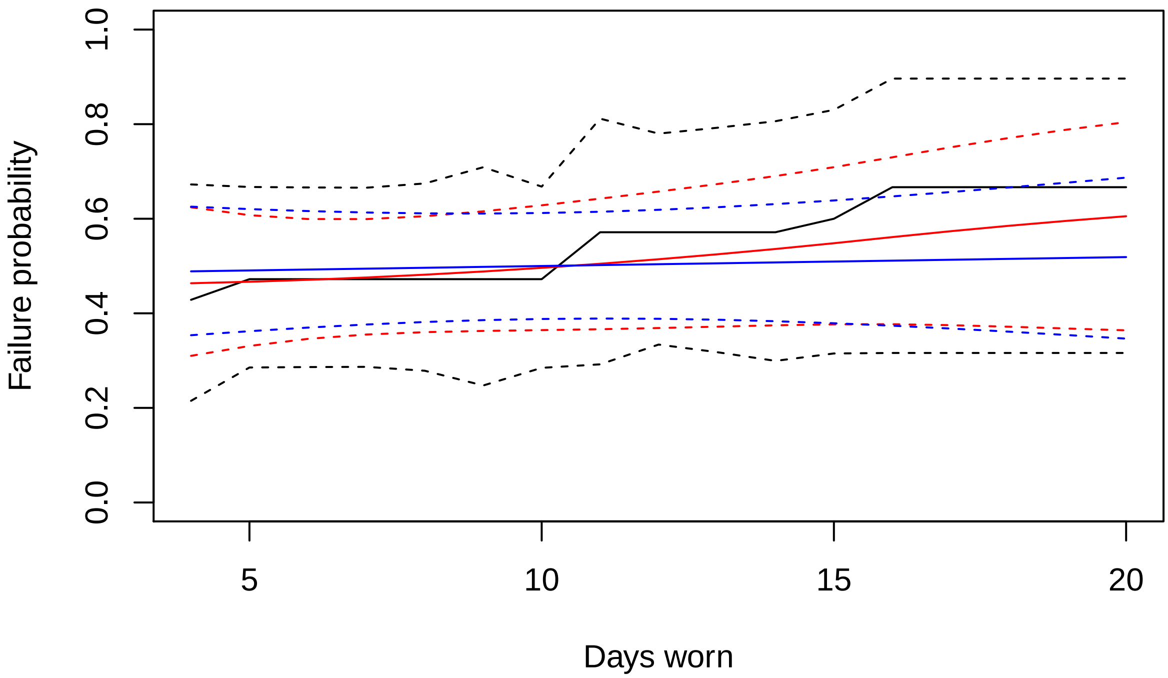

Figure 2: Comparison of cross sectional logistic regression models for mask Failure by number of days worn. Black = non-parametric monotone (step functions), Red = semi-parametric smooth with monotone constraint, Blue = semi-parametric smooth (no monotone requirement). Dashed lines 95% point-wise confidence limits. Fit with cgam package.

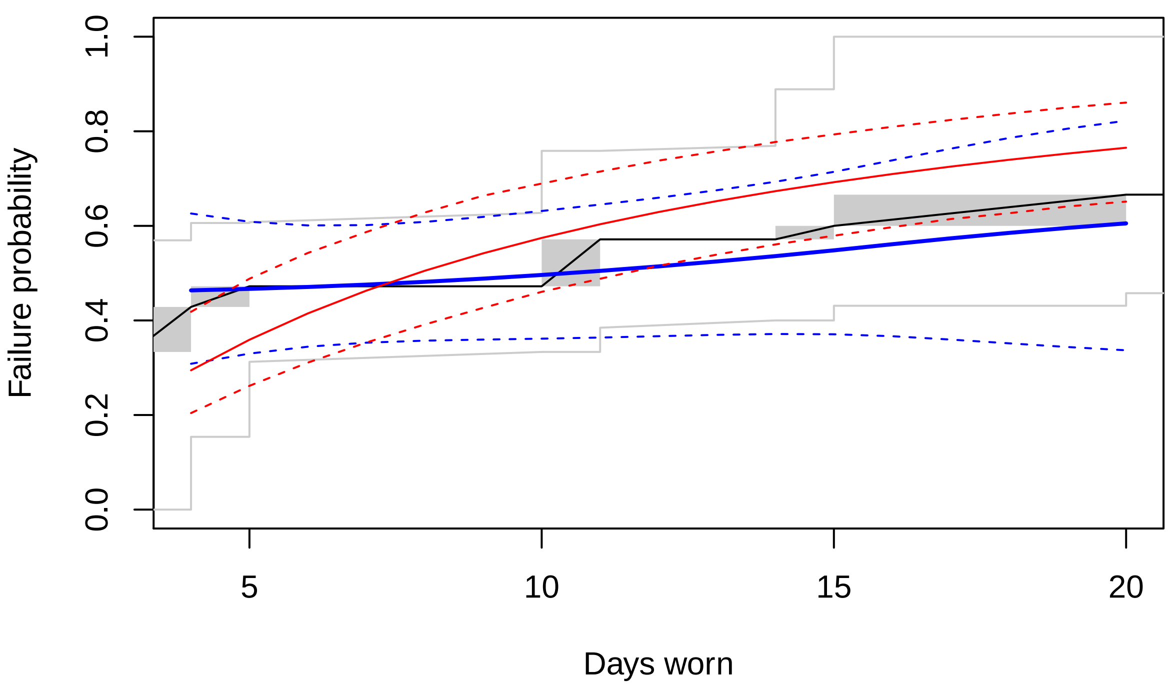

Figure 3: Comparison of survival and cross-sectional models for mask failure by number of days worn. Black = non-parametric survival with interval censoring (grey regions = indeterminate MLE), Red = generalized gamma survival model with interval censoring, Blue = cross sectional monotone smooth. Dashed lines 95% point-wise confidence limits.

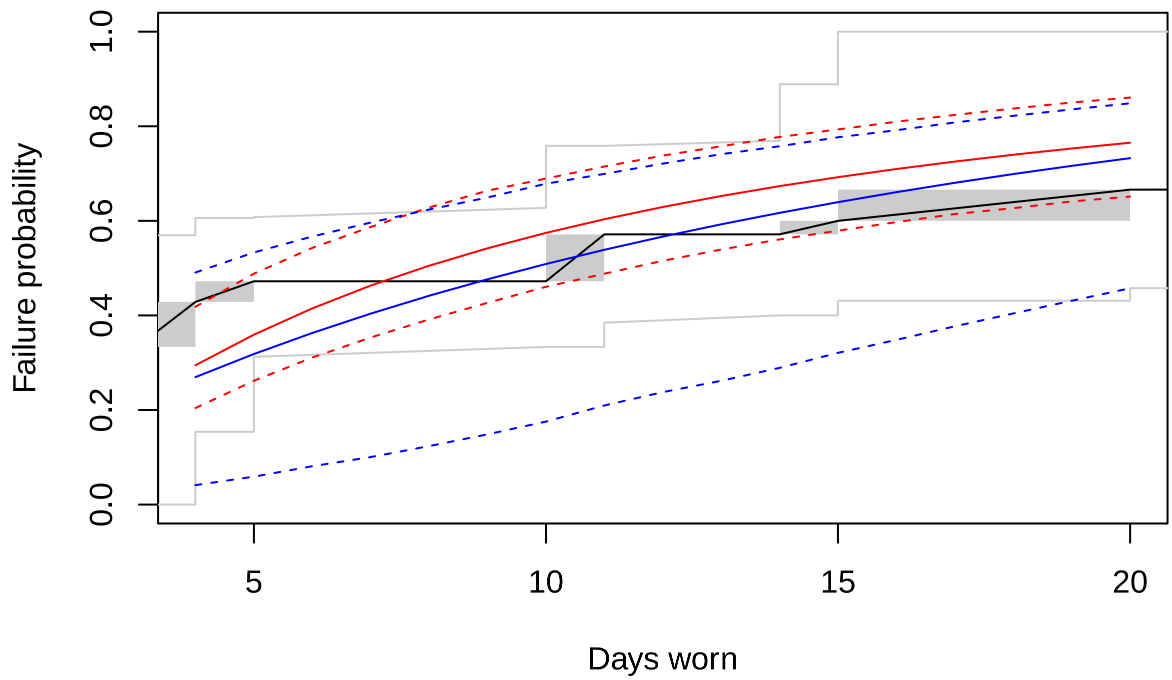

Figure 4: Comparison of parametric family survival models. Black = non-parametric survival (grey regions = indeterminate MLE), Red = generalized gamma survival, Blue = Weibull. Dashed lines 95% point-wise confidence limits. Fit with cgam package.

early failure.

Similar analysis for failure by number of times donned is shown in Figure 6. The failure fraction is similarly flat. Number of times used was fairly correlated with number of days worn (spearman correlation = 0.63, p < 0.001).

Alternative factors contributing to fit failure analyzed included number of sterilizations and intensity of use (number of times donned per day). Using the logistic regression model, the number of sterilizations was found to have a modest effect with modest precision (OR 1.5 95% CI 0.9, 2.8). The number of times donned per day was found to have a negligible effect with modest precision (OR 0.8 95% CI 0.6, 1). In cox analysis, the number of sterilizations was found to have a small effect with modest precision (cox B =0.16, 95% CI -0.23 to 0.55). The number of times donned per day was found to have a negligible effect with modest precision (cox B =-0.11, 95% CI -0.31 to 0.1).

Although imperfect, participants were able to somewhat discriminate poorly fitting masks. Figure 7 displays the survival stratified by perceived fit. Among the 12 participants reporting a poor fit, 83% failed the fit test (95% CI 55 to 95). Among the 62 participants reporting a good fit, 44% failed the fit test (95% CI 32 to 56, p 0.024 by Fisher's exact test). Reversing the direction of conditioning, among those passing the fit test, 5% believed their mask was poorly fitting (95% CI 1 to 18). Among those failing the fit test, 27% believed their mask to be poorly fitting (95% CI 15 to 43). The user's impression had a sensitivity of 27%, specificity of 95%, and positive and negative predictive values of 83% and 56%. Test administrators also judged mask quality. As seen in Table 1, very few masks were judged to be poor quality, but "Like New" and "Good" quality masks were 89% of the failed masks (95% CI 75 to 96). "Like New" masks were much less likely to fail, OR = 0.2 (95% CI =0.1 to 0.5, p < 0.001).

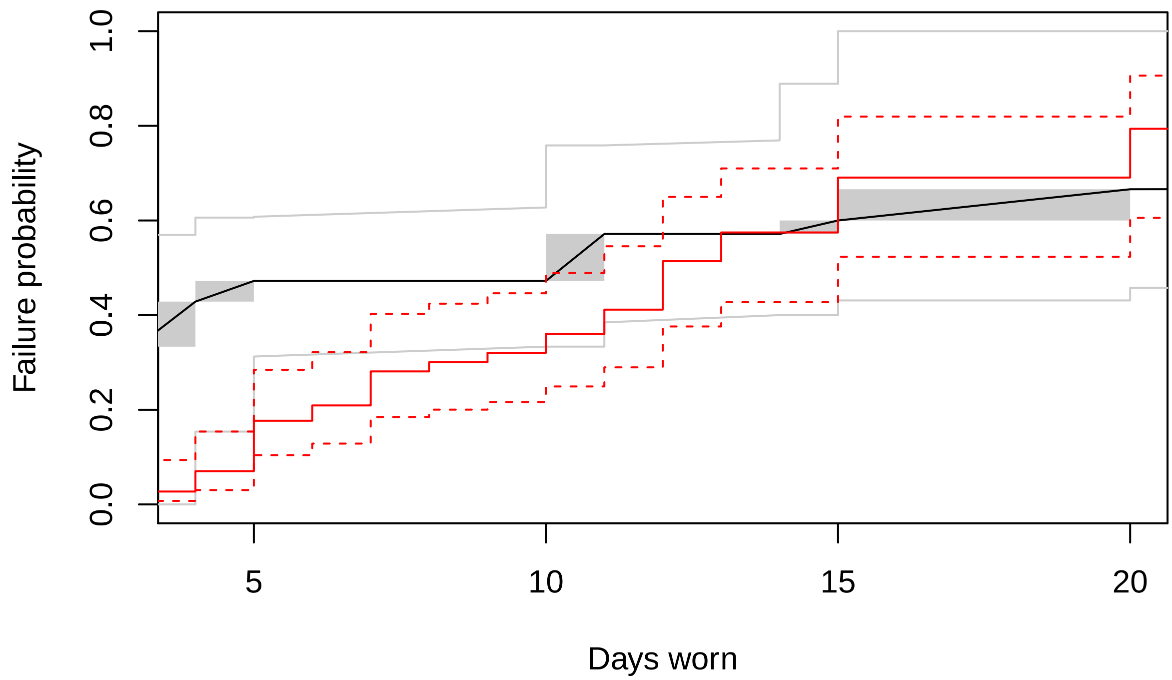

Figure 5: Comparison of interval censored and traditional survival estimates of mask failure by number of days worn. Black = non-parametric interval censored survival (grey regions = indeterminate MLE), Red = right censored only (Kaplan-Meier)

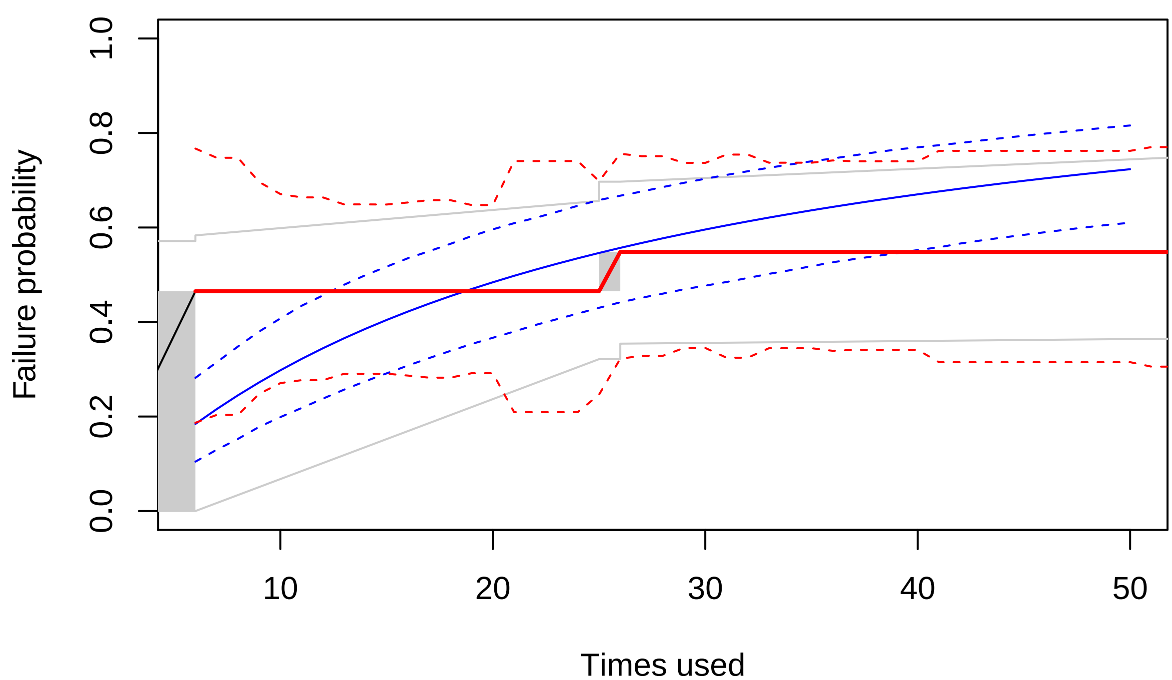

Figure 6: Comparison of survival and cross sectional models for mask failure by number of times used. Black = non-parametric interval censored survival (grey regions = indeterminate MLE), Red = cross-sectional analysis, Blue = generalized gamma survival

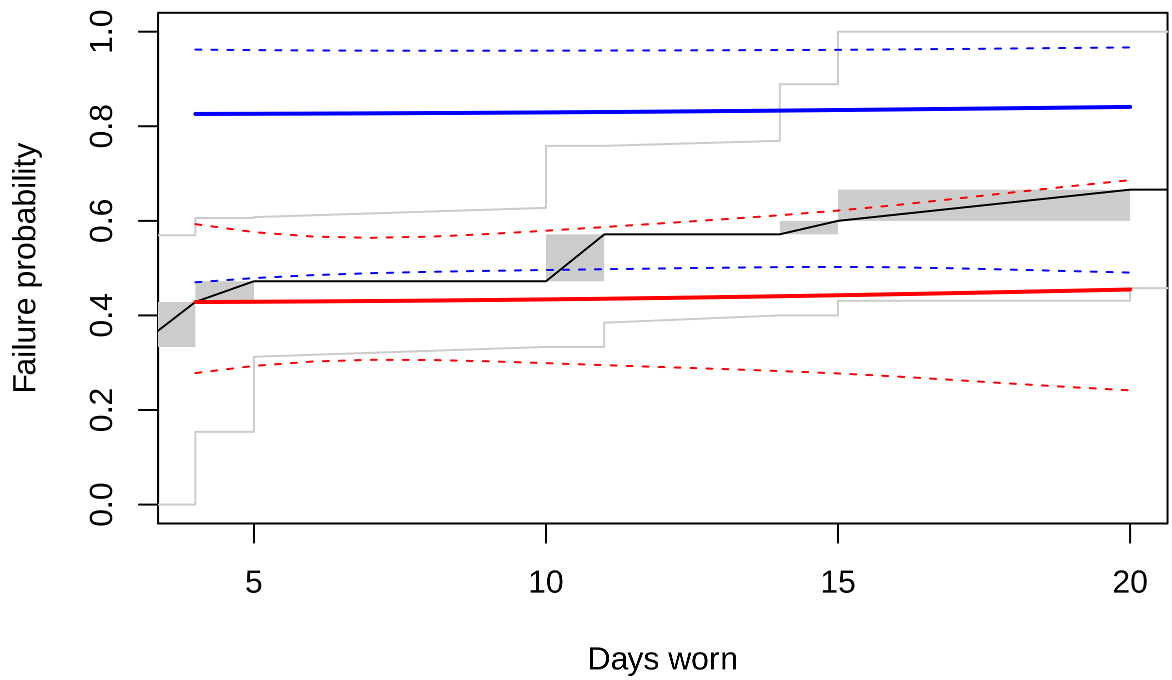

Figure 7: Comparison of mask failure by days worn stratified by user perceived fit. Black = entire study non-parametric interval censored survival (grey regions = indeterminate MLE), Red= smooth cross-sectional model among self-perceived good fit, Blue = smooth cross-sectional model among self-perceived bad fit

References

Anderson-Bergman, Clifford. 2017. "icenReg: Regression Models for Interval Censored Data in R." *Journal of Statistical Software* 81 (1): 1–23. https://doi.org/10.18637/jss.v081.i12.

Bouaziz, Olivier, Eva Lauridsen, and Grégory Nuel. 2018. "Regression Modelling of Interval Censored Data Based on the Adaptive Ridge Procedure." *arXiv:1812.09158 [Stat]*, December. http://arxiv.org/abs/1812.09158.

Groeneboom, P., and J. A. Wellner. 1992. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Oberwolfach Seminars. Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-8621-5.

Jackson, Christopher. 2016. "Flexsurv: A Platform for Parametric Survival Modeling in R." *Journal of Statistical Software* 70 (1): 1–33. https://doi.org/10.18637/jss.v070.i08.

Li, Chenxi, Daewoo Pak, and David Todem. 2019. "Adaptive Lasso for the Cox Regression with Interval Censored and Possibly Left Truncated Data:" *Statistical Methods in Medical Research*, June. https://doi.org/10.1177/0962280219856238.

Liao, Xiyue, and Mary C. Meyer. 2019. "Cgam: An R Package for the Constrained Generalized Additive Model." *Journal of Statistical Software* 89 (1): 1–24. https://doi.org/10.18637/jss.v089.i05.

Wang, Lianming, Christopher S. McMahan, Michael G. Hudgens, and Zaina P. Qureshi. 2016. "A Flexible, Computationally Efficient Method for Fitting the Proportional Hazards Model to Interval-Censored Data." *Biometrics* 72 (1): 222–31. https://doi.org/10.1111/biom.12389.

Zhao, Hui, Qiwei Wu, Gang Li, and Jianguo Sun. 2019. "Simultaneous Estimation and Variable Selection for Interval-Censored Data with Broken Adaptive Ridge Regression." *Journal of the American Statistical Association*, April. https://amstat.tandfonline.com/doi/abs/10.1080/01621459.2018.1537922.